

University of Groningen

## **Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network**

Qiu, Bingjiang; Guo, Jiapan; Kraeima, Joep; Glas, Haye H.; Borra, Ronald J. H.; Witjes, M. J. H.; van Ooijen, Peter M. A.

*Published in:*  
Physics in Medicine and Biology

*DOI:*  
[10.1088/1361-6560/ab2c95](https://doi.org/10.1088/1361-6560/ab2c95)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Final author's version (accepted by publisher, after peer review)

*Publication date:*  
2019

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Qiu, B., Guo, J., Kraeima, J., Glas, H. H., Borra, R. J. H., Witjes, M. J. H., & van Ooijen, P. M. A. (2019). Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network. *Physics in Medicine and Biology*, 64(17), [175020]. <https://doi.org/10.1088/1361-6560/ab2c95>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

ACCEPTED MANUSCRIPT

# Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network

To cite this article before publication: Bingjiang Qiu *et al* 2019 *Phys. Med. Biol.* in press <https://doi.org/10.1088/1361-6560/ab2c95>

## Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2019 Institute of Physics and Engineering in Medicine.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

# Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network

Bingjiang Qiu<sup>1,2</sup>, Jiapan Guo<sup>5</sup>, Joep Kraeima<sup>1,3</sup>, Haye H. Glas<sup>1,3</sup>, Ronald J. H. Borra<sup>2,4</sup>, Max J. H. Witjes<sup>1,3</sup> and Peter M. A. van Ooijen<sup>1,5</sup>

<sup>1</sup>3D Lab, <sup>2</sup>Department of Radiology, <sup>3</sup>Department of Oral and Maxillofacial Surgery, <sup>4</sup>Department of Nuclear Medicine and Molecular Imaging, <sup>5</sup>Department of Radiation Oncology  
University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ, Groningen, The Netherlands

E-mail: j.guo@umcg.nl

March 2019

**Abstract.** Segmentation of mandibular bone in CT scans is crucial for 3D virtual surgical planning of craniofacial tumor resection and free flap reconstruction of the resection defect, in order to obtain a detailed surface representation of the bones. A major drawback of most existing mandibular segmentation methods is that they require a large amount of expert knowledge for manual or partially automatic segmentation. In fact, due to the lack of experienced doctors and experts, high quality expert knowledge is hard to achieve in practice. Furthermore, segmentation of mandibles in CT scans is influenced seriously by metal artifacts and large variations in their shape and size among individuals. In order to address these challenges we propose an automatic mandible segmentation approach in CT scans, which considers the continuum of anatomical structures through different planes. The approach adopts the architecture of the U-Net and then combines the resulting 2D segmentations from three orthogonal planes into a 3D segmentation. We implement such a segmentation approach on two head and neck datasets and then evaluate the performance. Experimental results show that our proposed approach for mandible segmentation in CT scans exhibits high accuracy.

Keywords: automatic mandible segmentation, convolutional neural network (CNN), 3D virtual surgical planning, oral and maxillofacial surgery

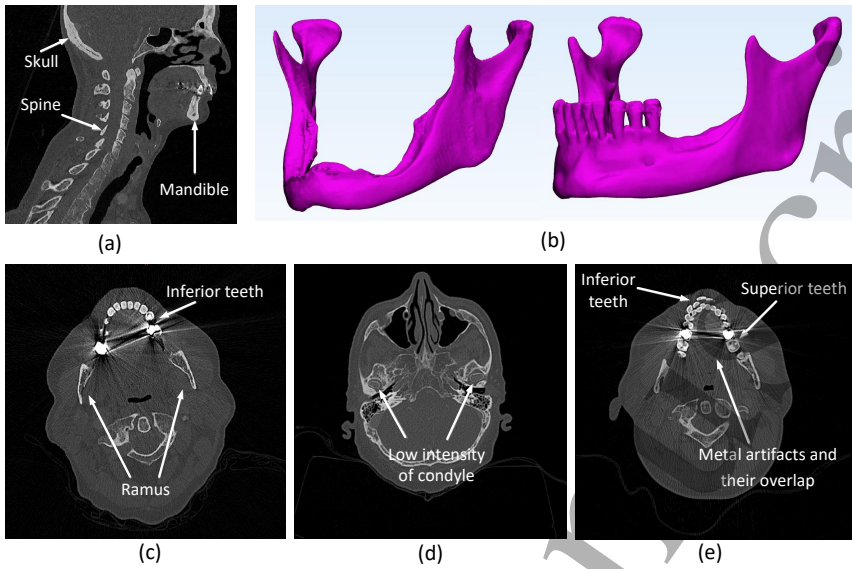
## 1. Introduction

Three-dimensional virtual surgical planning (3D VSP) is proposed to be a precise and predictable method for bone related craniofacial tumor resection and free flap reconstruction of the mandible (Bittermann, Scheifele, Prokic, Bhatt, Henke, Grosu,

Schmelzeisen & Metzger 2013)(Essig, Rana, Kokemueller, von See, Ruecker, Tavassol & Gellrich 2011)(Schepers, Raghoobar, Vissink, Stenekes, Kraeima, Roodenburg, Reintsema & Witjes 2015)(Weijs, Coppen, Schreurs, Vreeken, Verhulst, Merckx, Bergé & Maal 2016)(Kraeima, Schepers, van Ooijen, Steenbakkers, Roodenburg & Witjes 2015). It is performed pre-operatively to determine the resection margins and osteotomy planes. The planning is translated towards the actual surgical procedure by the use of patient specific 3D printed guides. Currently, computed tomography (CT) is the most commonly used modality to implement such a process. The segmentation of anatomical structures plays a critical role in the 3D VSP. Conventional manual segmentation of the mandible in CT scans leads to a tedious procedure in the clinical practice (Huff, Ludwig & Zuniga 2018). Moreover, the structural complexity of mandibles and the considerable human rater variability make the segmentation of mandibles in CT scans challenging (Torosdagli, Liberton, Verma, Sincan, Lee, Pattanaik & Bagci 2017). Manual segmentation also has limited reproducibility and is very time-consuming. Therefore, semiautomatic or automatic image segmentation would improve efficiency and reliability, as well as reduce the workload of technologists (Huff et al. 2018).

Automatic and accurate mandible segmentation remains challenging for several reasons. First, head and neck CT scans cover various bone-structured organs of the human body, such as the skull and the spine, Figure 1(a). Second, there is a large variation in the appearance of anatomical structures of mandibles, Figure 1(b). Third, noise and other artifacts from the teeth (e.g. because of fillings or braces) as well as lower intensity in the condyles very often lead to ambiguous and blurred boundaries in CT scans, Figure 1(c-d). Fourth, inconsistent manual labeling can confuse the trained models for mandible segmentation. For instance, the superior teeth are labeled as part of the mandible while inferior teeth are not, and the presence of both the superior and the inferior teeth in the same slice (Torosdagli et al. 2017), Figure 1(e). All these factors make the automatic segmentation of mandibles difficult by using mathematical models that are applicable to various imaging cases.

Conventional approaches for mandible segmentation typically use pixel-based or model-based methods. In the past decade, some traditional semi-automatic and automatic methods have been developed to segment the mandible in CT scans (Gollmer & Buzug 2012)(Torosdagli et al. 2017)(Chuang, Doherty, Adluru, Chung & Vorperian 2017)(Abdi, Kasaei & Mehdizadeh 2015)(Chen & Dawant 2015)(Mannion-Haworth, Bowes, Ashman, Guillard, Brett & Vincent 2015)(Albrecht, Gass, Langguth & Lüthi 2015). A statistical shape model for mandible segmentation was presented by Gollmer and Buzug in 2012 (Gollmer & Buzug 2012). Torosdagli et al. proposed a 3D gradient-based fuzzy connectedness algorithm to segment mandibles (Torosdagli et al. 2017). A registration-based semiautomatic mandible segmentation technique was proposed by Chuang et al. in 2017 (Chuang et al. 2017). The work of Abdi et al. applies an automatic segmentation algorithm via collecting superior, inferior and exterior borders of mandibles in panoramic x-rays (Abdi et al. 2015). Chen et al. (Chen & Dawant 2015) proposed a multi-atlas model that registered CT images



**Figure 1.** Examples of typical cases that challenge the accurate mandible segmentation in CT scans. (a) Various bone-structured organs in the head and neck CT, such as the skull and the spine. (b) Large variation of mandibles between individuals. (c) Metal artifacts and noises in the teeth. (d) Lower intensity in the condyles. (e) Metal artifacts and the presence of inferior and superior teeth in the same slice.

with the atlases at the global level, which allowed multi-atlas-based segmentations and correlation-based label fusion to be performed at the local level. The method from (Mannion-Haworth et al. 2015) applied active appearance models (AAM) built from manually segmented examples. High quality anatomical correspondences for the models are generated using a groupwise registration method. The models are then applied to segmentation of ROIs in CT scans. Albrecht et al. employed a multi-atlas segmentation to obtain an initial segmentation and then apply an Active Shape Model (ASM) segmentation to refine the initial segmentation of the organ (Albrecht et al. 2015). The performances of the conventional methods, however, are often affected by the noise or metal artifacts in the CT images. Weak and false edges in the parts of condyles and teeth, often appear in the detected images, which frustrate the accurate segmentation of the mandible. The proposed deformable models, of which the parameters are determined according to the global characteristics of the target contour, are difficult to adapt to some local areas of the contour. (Yuheng & Hao 2017)(Blaschke, Burnett & Pekkarinen 2004)(Bankman 2008)

Since 2013, convolutional neural networks (CNNs) have been successfully applied in computer vision tasks, for example, image classification(Krizhevsky, Sutskever & Hinton 2012), super-resolution(Dong, Loy, He & Tang 2014), medical imaging(He, Yang, Wang, Zeng, Bian, Zhang, Sun, Xu & Ma 2019)(Li, Zeng, Peng, Bian, Zhang, Xie, Wang, Liao, Zhang, Huang et al. 2019) and semantic segmentation(Long, Shelhamer &

*Phys. Med. Biol.*

4

Darrell 2015). Medical image segmentation, as one of the research focuses in semantic segmentation, has developed exponentially due to the rapid evolution of CNNs (Ker, Wang, Rao & Lim 2018)(Shen, Wu & Suk 2017). Semantic segmentation aims to assign a semantic label at the pixel level (Fu 2012). Recent advances in semantic segmentation (Long et al. 2015)(Badrinarayanan, Kendall & Cipolla 2017)(Ronneberger, Fischer & Brox 2015)(Yu & Koltun 2015)(Chen, Papandreou, Kokkinos, Murphy & Yuille 2018)(Lin, Milan, Shen & Reid 2017)(Zhao, Shi, Qi, Wang & Jia 2017)(Peng, Zhang, Yu, Luo & Sun 2017)(Chen, Papandreou, Schroff & Adam 2017)(Garcia-Garcia, Orts-Escolano, Oprea, Villena-Martinez & Garcia-Rodriguez 2017)(Qin, Wu, Han, Yuan, Zhao, Ibragimov, Gu & Xing 2018)(Wu, Tha, Xing & Li 2018) have enabled their applications to medical image segmentation. Most of the above mentioned CNN architectures have been proven to be effective on semantic segmentation of natural scene images. After that, Ibragimo et al. (Ibragimov & Xing 2017) presented the first attempt of using deep learning concept of CNN to segment organs-at-risks in head and neck CT scans. The AnatomyNet (Zhu, Huang, Tang, Qian, Du, Fan & Xie 2018) is built upon the popular 3D U-net architecture using residual blocks in encoding layers and a new loss function combining Dice score and focal loss in the training process. A fully CNN (FCNN) method was presented by (Tong, Gou, Yang, Ruan & Sheng 2018). The multi-planar training strategy was presented by (Mortazi, Burt & Bagci 2017). Such works motivate our implementation of CNNs for automatic mandible segmentation in CT scans.

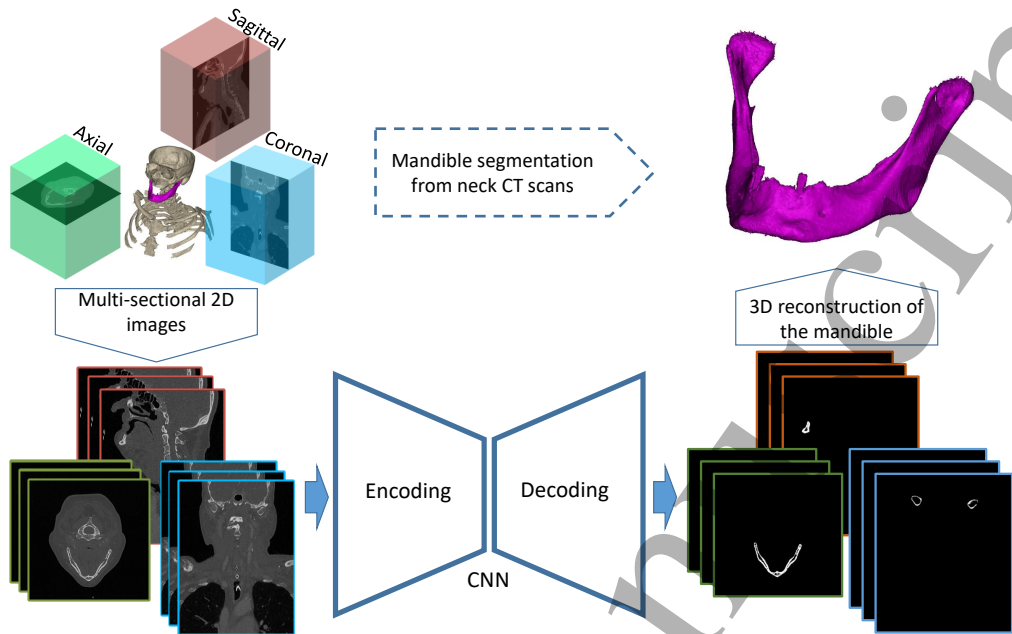
Here, we propose a CNN-based approach for 3D mandible segmentation in CT scans. Such an approach takes a multi-planar volume-to-slice strategy which takes into account the spatial information of adjacent slices in order to preserve the connectivity of anatomical structures.

## 2. Materials and methods

### 2.1. Data preparation

The collection of the patient data sets for medical research purposes was approved by the local Medical Ethical Committee. The data set contains 109 CT scans reconstructed with a kernel of Br64, I70h(s) or B70s. Each scan consists of 221 to 955 slices with a size of  $512 \times 512$  pixels. The pixel spacing varies from 0.35 to 0.66 mm, and the slice thickness varies from 0.6 to 0.75 mm. The manual mandible segmentation was performed using Mimics software version 20.0 (Materialise, Leuven, Belgium) by a trained researcher and confirmed by a clinician.

We also validate our method on a public dataset of head and neck scans dataset, which was obtained from Public Domain Database for Computational Anatomy (PDDCA) version 1.4.1 (Raudaschl, Zaffino, Sharp, Spadea, Chen, Dawant, Albrecht, Gass, Langguth, Lüthi et al. 2017). The original CT dataset is provided and maintained by Dr. Gregory C Sharp (Harvard Medical School – MGH, Boston) and his group.



**Figure 2.** The proposed framework of the multi-planar volume-to-slice network for 3D segmentation of the mandible.

PDDCA version 1.4.1 comprises 48 patient CT images from the Radiation Therapy Oncology Group (RTOG) 0522 study, a multi-institutional clinical trial, together with manual segmentation of left and right parotid glands, brainstem, optic chiasm and mandible. Each scan consists of 76 to 360 slices with a size of  $512 \times 512$  pixels. The pixel spacing varies from 0.76 to 1.27mm, and the slice thickness varies from 1.25 to 3.0 mm. Forty of the 48 patients in PDDCA with manual mandible annotation are used in this study. (Raudaschl et al. 2017)(Ren, Xiang, Nie, Shao, Zhang, Shen & Wang 2018)

2.2. Methods

We illustrate our proposed framework for automatic 3D mandible segmentation in Figure 2. Such a framework takes as input the 2D slice images of the head and neck CT scans from three orthogonal (axial, coronal and sagittal) planes. The images from each plane are fed into a convolutional neural network that automatically segments the mandible. We then combine the 2D segmentation results obtained from each plane to reconstruct into a 3D segmentation of the mandible.

Our decision to combine the 2D segmentation results from three orthogonal planes for 3D mandible segmentation is motivated by two reasons. First, the anatomical structures of mandibles can be more representative from multi-planar images which are then used to train the computational models for mandible segmentation, comparing to the use of single plane images. Second, it helps to reduce computation complexity in comparison to direct 3D segmentation on volumetric CT scans.

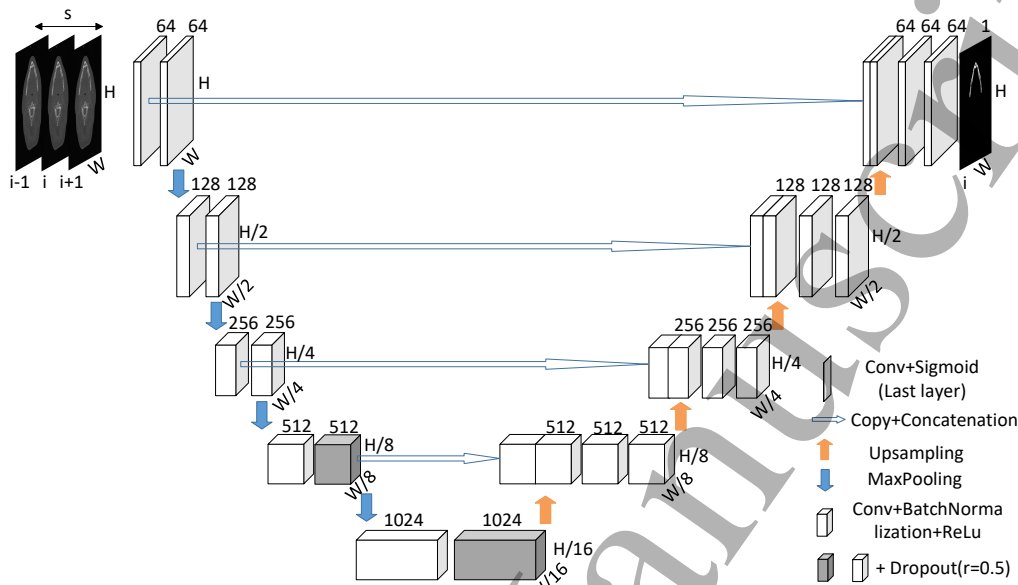
**2.2.1. Single-planar CNN model** In this subsection, we elaborate on the design of the single-planar module. Figure 3 shows the architecture of the single-planar module, which is inspired by that of the U-Net (Ronneberger et al. 2015). In order to consider the similarity and continuity of upper and lower regions of the mandible, we use multi-sectional slices for training in each plane, which can help retain the structural information of the mandible. We call this strategy multi-planar volume-to-slice. This single-planar network consists of an encoding and a decoding procedure of in total 23 convolutional layers, each of which has convolutional kernels of the same size  $3 \times 3$ . The number of feature maps is listed on the top of each block that represents the convolutional layer in Figure 3. During the encoder procedure, max pooling layers are used to further enlarge the receptive fields (LeCun, Bengio & Hinton 2015). In the configuration of the model, we increase the number of feature maps by a factor of 2 after every max pooling layer. For the technical details on the architecture of the U-Net, we refer the interested readers to (Ronneberger et al. 2015). There are two dropout layers (Srivastava, Hinton, Krizhevsky, Sutskever & Salakhutdinov 2014) with a dropout rate  $r = 0.5$  in the encoder, as shown by the grey blocks in Figure 3. Dropout is a technique to address the issues of overfitting that usually makes CNN underperform during the test phase. The mechanism of dropout is to randomly drop units along with their connections during the training of CNNs so that it exhibits the bagged ensemble of many neural networks (LeCun et al. 2015). In such a way it helps to prevent CNN from overfitting in a computationally inexpensive but powerful manner (Srivastava et al. 2014). During the decoder procedure, upsampling layers are used and the feature maps are then concatenated with those of the same resolution from the encoder procedure. Such feature concatenation is indicated by the horizontal arrows in Figure 3. In the original U-Net, the resulting feature maps from each level in the encoder are concatenated to those on the same level in the decoder. In order to save computational complexity, we concatenate the feature maps from the first convolutional layer (instead of the second convolutional layer in each level) of the level in the encoder with those in the decoder, which is indicated by the horizontal arrows in Figure 3.

Each of the convolutional layer blocks is composed of a linear convolution, a batch normalization (Ioffe & Szegedy 2015) and an element-wise nonlinear ReLU function (Nair & Hinton 2010), in which batch normalization is a technique for improving the performance and stability of CNN. Finally, the resulting responses are turned into probability values using a sigmoid classifier (Lin & Lin 2003). The output has the same size as that of the input image, as shown in Figure 3. In general, the convolutional layers (illustrated by the cubic blocks in Figure 3) can be expressed in the following equation:

$$L_i = \Gamma(L_{i-1} \otimes K_{ij} + B_i), \quad (1)$$

where  $L_i$  is the output of the  $i$ -th layer ( $i \in [1, 23]$ ) and the number of feature maps for  $L_i$  is shown in the top of each layer in Figure 3,  $\Gamma$  represents the nonlinear operator of ReLU or sigmoid function in the output or last layer,  $\otimes$  is the convolutional operator,





**Figure 3.** The 2D single-planar mandible segmentation based on the U-Net architecture. In the input layer, we denote the height, the width and the number of slices by  $H$ ,  $W$  and  $s$ , respectively. All convolutional kernels have a size of  $3 \times 3$ .

$K_{ij}$  represents the  $j$ -th feature kernel in the  $i$ -th layer and  $B_i$  denotes a bias of the  $i$ -th layer.

We use a loss function based on Dice coefficient which is commonly used to evaluate the performance of image segmentation tasks (Ghafoorian, Karssemeijer, Heskens, Uden, Sanchez, Litjens, Leeuw, Ginneken, Marchiori & Platel 2017). We provide the detailed information about Dice Coefficient loss in Section 2.2.3.

It is worth noting that the U-Net is able to predict the probability of scores (in our case the probability of the mandible) based on the structural texture information within a given receptive field. This enables the use of all information in a 2D CT plane to directly predict complex structures. The U-Net architecture provides the capability to adapt naturally to input CT images of any resolution.

**2.2.2. Multi-planar mandible segmentation framework** The use of consecutive slices from the scans helps improve the 2D segmentation of the mandible. It, however, does not provide satisfactory results since some parts of the mandible shapes are better observed via other planes. Thus, we create a three-planar network which takes into account the structural information of the mandible from different orthogonal planes. Such a multi-planar network consists of three subnetworks, each of which is fed with

*Phys. Med. Biol.*

8

CT data from one of the three planes. In the network fed with data from the axial plane, the size of the input data has height  $H = 512$ , width  $W = 512$  and number of slices  $s = 3$ . In the Sagittal (Coronal) planes which often have different numbers of slices in every scan, we use a sliding window to crop the input images into the same size of  $H = 400$ ,  $W = 400$  and  $s = 7$  ( $H = 400$ ,  $W = 400$  and  $s = 9$ ). We use Adam optimization (Kingma & Ba 2014) with a learning rate of  $10^{-5}$ . The three subnetworks use the same size of convolution kernels of  $3 \times 3$  with zero padding.

*2.2.3. Loss metric* The Dice similarity index, also called the Dice score, is often used to measure consistency between two samples (Ghafoorian et al. 2017). Therefore, it is widely applied as a metric to evaluate the performance of image segmentation algorithms. It is defined as:

$$\text{Dice} = \frac{2|Y_r \cap Y_p|}{|Y_r| + |Y_p|}, \quad (2)$$

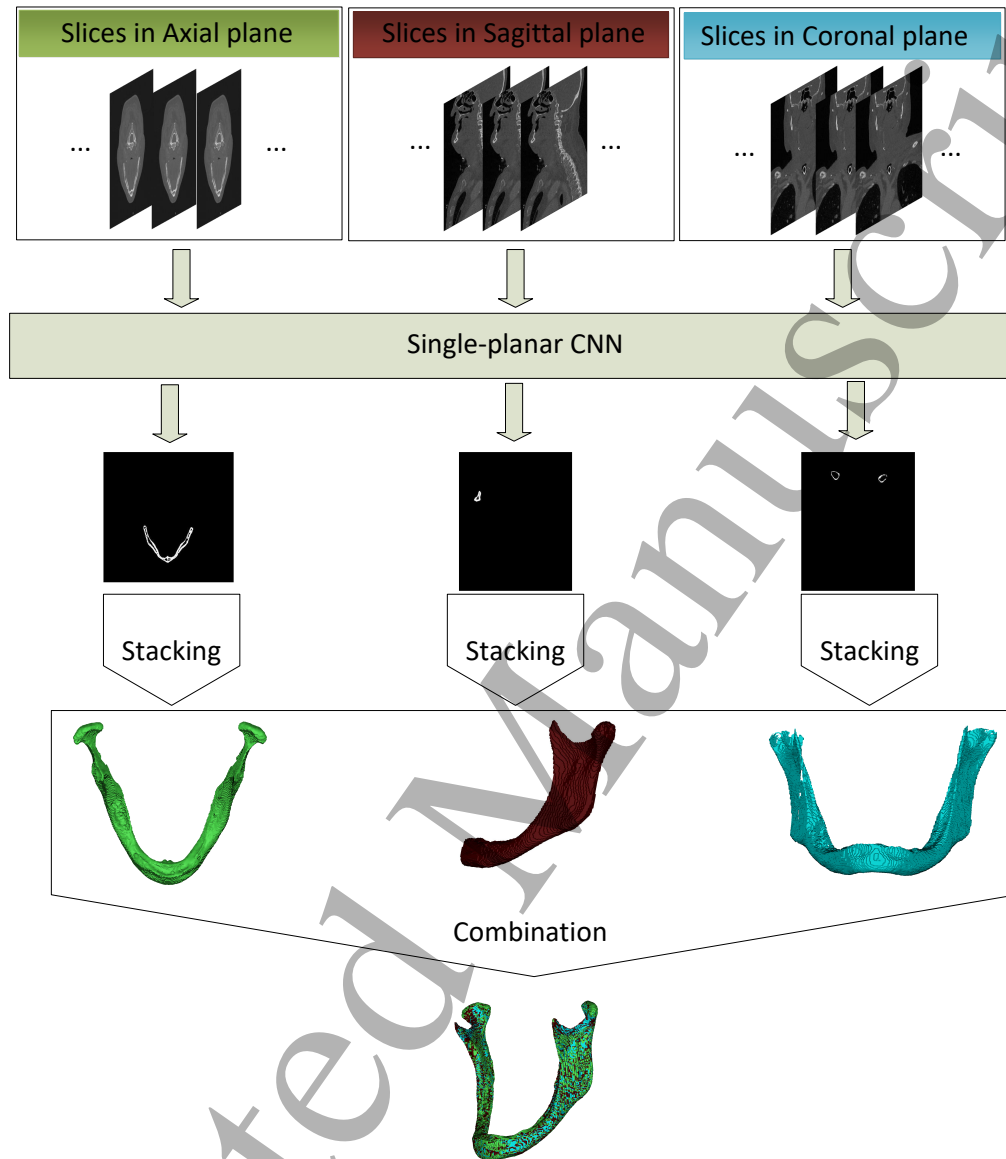
where  $Y_r$  is the reference standard, and  $Y_p$  is the prediction from the network. Such a score has a value between 0 and 1, in which 0 means the total disagreement between the reference standard and the evaluated segmentation while 1 presents total agreement. A differentiable formulation of the Dice coefficient has been proposed by Milletari (Milletari, Navab & Ahmadi 2016), for minimizing the loss ( $\text{loss} = 1 - \text{Dice}$ ) between the two binary labels in the training of the proposed multi-planar network.

*2.2.4. Combination of the multi-planar network for 3D mandible segmentation* We train each of the subnetworks in the proposed multi-planar framework independently and then apply the trained framework on the test scans. In order to obtain the 3D segmentation of the mandible, we stack the 2D segmentation results from each subnetwork into a 3D volume data, in which each voxel has a probability prediction from the sigmoid function in each subnetwork. For instance, the resulting output (probability prediction) from three subnetworks can be denoted by  $Y^a$ ,  $Y^s$  and  $Y^c$  separately. In order to effectively combine the predicted information, we select the maximum probability of each voxel from the three networks,  $Y = \max\{Y^a, Y^s, Y^c\}$ , to improve accuracy of the final results (Mortazi et al. 2017). Figure 4 shows the combination procedures.

### 3. Experiments

#### 3.1. Evaluation criterion

We use Dice similarity coefficient and the 3D surface error to evaluate the performance of the proposed approach for mandible segmentation. We compute the Dice similarity coefficient of the automatic segmentation results with respect to the manual segmentation on both 2D slice images and the complete 3D volumetric data. In order to observe the segmentation results in a more straightforward way, we use the Materialise

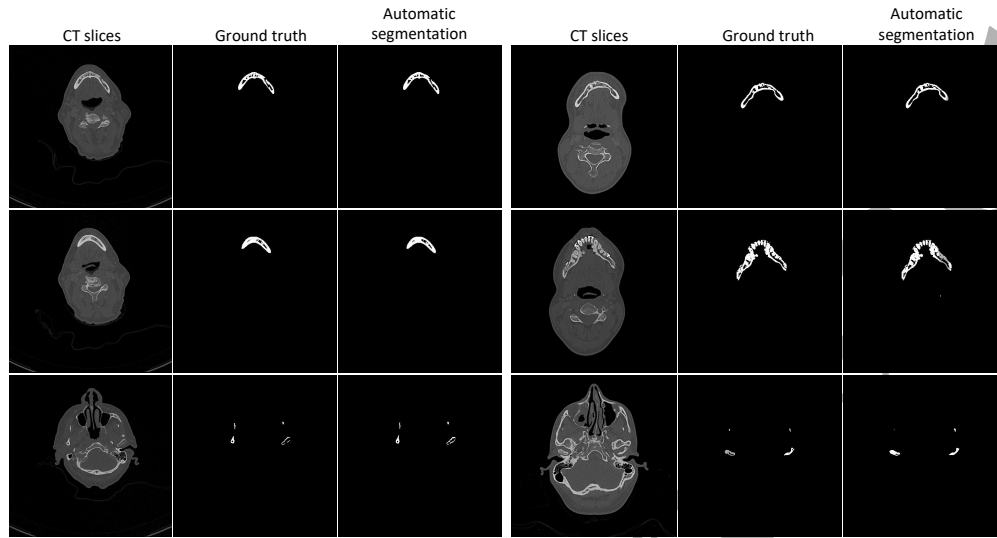


**Figure 4.** Step-by-step illustration on the 3D mandible segmentation.

Mimics software to reconstruct the 2D automatic segmentation into a 3D view. We then use the Materialise 3-matic software to automatically post-process the segmentation results in order to remove the disconnected voxels. Afterwards, we compute the root mean square error (RMS) of the mandible surfaces between the manual and the post-processed model.

### 3.2. Experimental results

We implement the framework proposed in Section 2 by using Keras (Chollet et al. 2015) package with the Tensorflow (Abadi, Agarwal, Barham, Brevdo, Chen, Citro, Corrado, Davis, Dean, Devin et al. 2016) backend. The CNN models are trained on a workstation



**Figure 5.** Examples of the automatic segmentation of mandibles in the axial plane.

equipped with Nvidia Tesla K40m GPU of 12GB memory.

**3.2.1. Experiments on 109 CT scans** We randomly chose 52 cases as training, 8 cases as validation and 49 cases as test. We repeated the above mentioned process three times and then evaluated the performance of the proposed approach for automatic mandible segmentation. The training of a single-plane model takes less than 60 hours while the duration of the test on one scan is about 10 minutes. Figure 5 illustrates several examples of the results achieved in the axial direction by the proposed approach on the test scan. The average Dice score on all test scans from three repeats are 0.893, 0.878 and 0.872, respectively, as listed in Table 1.

**Table 1.** Dice score and RMS surface error on the three random repeats of the experiments.

	Test 1	Test 2	Test 3	Average
Dice score on the axial plane	0.799	0.767	0.789	0.785
Dice score on the sagittal plane	0.559	0.519	0.391	0.490
Dice score on the coronal plane	0.594	0.531	0.545	0.557
Dice score of the combined results	0.893	0.878	0.872	0.881
RMS value (mm) of the combined results	0.4284	0.5349	0.7740	0.5791

Figure 6 shows several cases of the postprocessed automatic segmentation and their comparison with the manual annotation in the 3D view. The visual comparison of the automatic segmentation with manual segmentation illustrates the effectiveness of our method on automatic segmentation of mandible.

Figure 7-9 illustrates the distribution of the root mean square errors of the surface distances, in which the blue lines indicate the average RMS values that is also listed in Table 1. From Figure 7-9, the majority of the test cases have the surface error under the average errors. We also provide the visual 3D illustration on the cases that have the top three RMS surface errors in each repeat. One of the main reasons for the large surface error is that parts of superior teeth are also automatically segmented while only the inferior teeth are manually annotated.

*3.2.2. Experiments on the PDDCA dataset* In particular, we compare our proposed method with several state-of-the-art methods on PDDCA dataset. We use the model based on test 1, i.e., we use the weights trained in test 1 of section 3.2.1 as initialization parameters. According to the Challenge description, we follow the same training and testing protocol (Raudaschl et al. 2017). A subset of 40 scans was used: 25 scans (0522c0001-0522c0328) are used as training data, and 15 scans (0522c0555-0522c0878) are used for testing (Raudaschl et al. 2017).

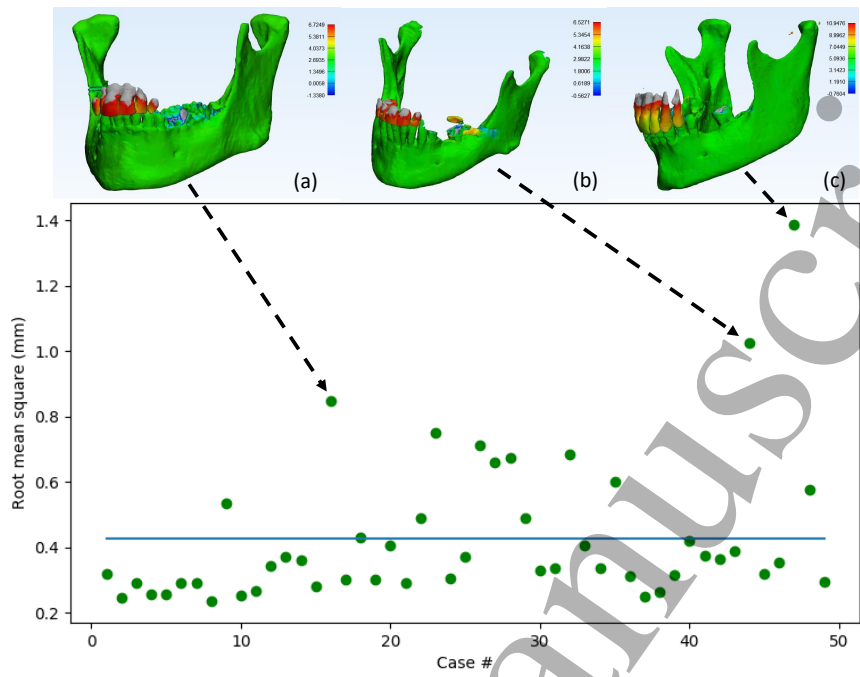
For comparison purpose, Table 2 lists Dice score and 95% Hausdorff distance (95HD) used in the Challenge paper (Raudaschl et al. 2017)(Tong et al. 2018). According to Table 2, the proposed method outperforms most other methods, with the second highest mean DC score and the lowest 95HD. For Dice score results, the segmentation result of our method is only slightly worse than the FCNN+SRM(Tong et al. 2018), while outperforming the rest of the works.

**Table 2.** Comparison of average Dice ( $\pm$ standard deviation) between the state-of-the-art methods and our method, bold fonts indicate the best performer for that structure.

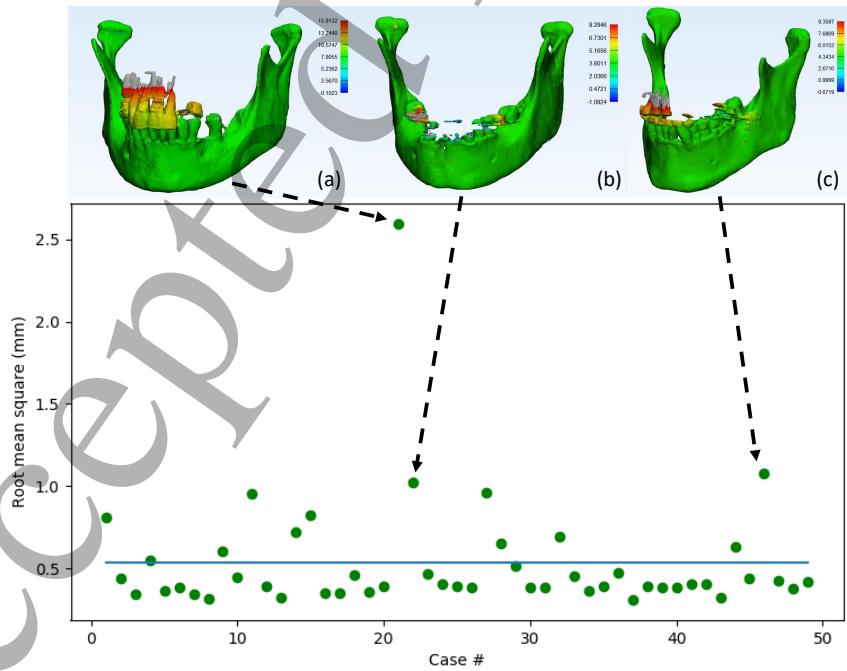
Methods	Dice	95HD (mm)
Multi-atlas(Chen & Dawant 2015)	0.917( $\pm$ 0.0234)	2.4887( $\pm$ 0.7610)
AAM(Mannion-Haworth et al. 2015)	0.9267	1.9767
ASM(Albrecht et al. 2015)	0.8813( $\pm$ 0.0555)	2.832( $\pm$ 1.1772)
CNN(Ibragimov & Xing 2017)	0.895( $\pm$ 0.036)	-
AnatomyNet(Zhu et al. 2018)	0.9251( $\pm$ 0.02)	(6.28 $\pm$ 2.21)
FCNN(Tong et al. 2018)	0.9207( $\pm$ 0.0115)	2.01( $\pm$ 0.83)
FCNN+SRM(Tong et al. 2018)	<b>0.936</b> ( $\pm$ 0.0121)	1.5( $\pm$ 0.32)
The proposed method	0.9328( $\pm$ 0.0144)	<b>1.4333</b> ( $\pm$ 0.5564)



**Figure 6.** Visual examples of final segmentations on our data set. From left to right are 3D views of (a) ground truth, (b) the algorithm segmentation, (c) the surface distance maps (mm) from algorithm segmentation to manual segmentation.

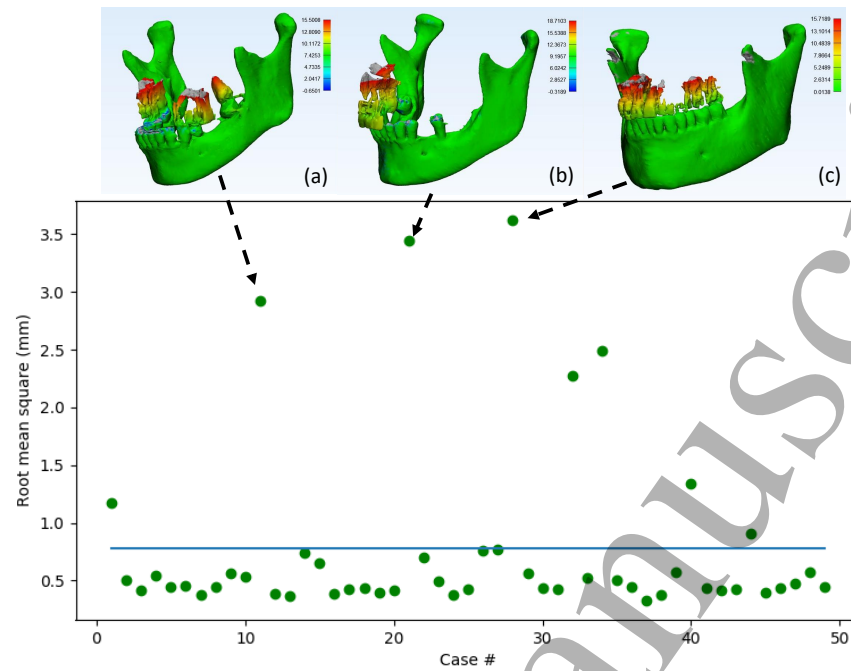


**Figure 7.** Test 1: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS vales in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 0.8494 mm, 1.0261 mm and 1.3885 mm



**Figure 8.** Test 2: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS vales in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 2.5965 mm, 1.0261 mm and 1.0744 mm





**Figure 9.** Test 3: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS values in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 2.9267 mm, 3.4411 mm and 3.6207 mm

#### 4. Discussion

In this work, we present a CNN-based method for mandible segmentation that is demonstrated to be effective for 3D segmentation of the mandible in a data set of 109 head and neck scans. Our approach uses multi-sectional CT slices from different orthogonal planes as input and then combines the 2D segmentation results from each plane in order to achieve a mandible segmentation in a 3D view. The proposed approach spends around 2.5 minutes to process one scan in the PDDCA dataset. Table 3 compares the testing time to segment a new patient using our method versus the state-of-the-art methods in the PDDCA dataset. In particular, the testing time costs increase to 10 minutes in the 109 CT scans. We obtain similar accuracies compared to the work of the traditional methods (Mannion-Haworth et al. 2015)(Torosdagli et al. 2017) but with superiority in efficiency and fully-automatic nature. Furthermore, such a design of the system enables the consideration of similarity and structural continuity of the mandible from different planes. The main contributions of this work can be summarized as follows.

First, the proposed approach extracts discriminative features for segmentation from three orthogonal planes (axial, coronal and sagittal) since some parts of the mandible shapes are better observed via other planes. In our study, we make full use of the extracted information from different planes to segment the mandible. Our proposed method indicates that 3D segmentation of medical images can also be achieved by 2D segmentation network which could help alleviate memory issues. Remarkably, this



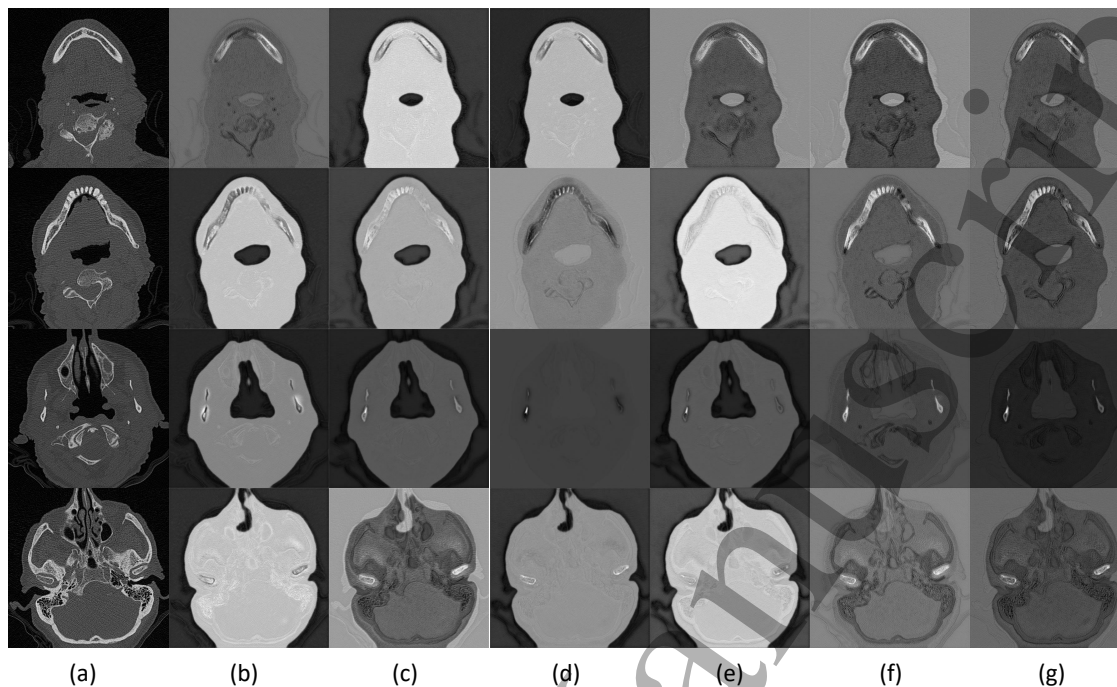
**Table 3.** Comparison of segmentation time on PDDCA dataset between the state-of-the-art methods and our method.

Methods	Segmentation time	Experimental equipment
Multi-atlas(Chen & Dawant 2015)	Over 60 min per patient	CPU
AAM(Mannion-Haworth et al. 2015)	30 min per image	CPU
ASM(Albrecht et al. 2015)	24 min per patient (in the online Challenge)	CPU
CNN(Ibragimov & Xing 2017)	30 min per patient	GPU
AnatomyNet(Zhu et al. 2018)	0.12 sec per patient	GPU
FCNN and FCNN+SRM(Tong et al. 2018)	9.50 sec per patient	GPU
The proposed method	2.5 min per patient	GPU

method has a good performance on mandible segmentation. For example, Figure 10 shows the intermediate feature maps in the 21-st layer from the Axial plane, which illustrates that the neural network actually learned the structural representation of mandibles. The outputs in Figure 10 mainly focus on the mandible; even the parts of condyles and ramus are accurately marked out. The intermediate results demonstrate the feasibility of extracting image features by the proposed CNN-based framework. Moreover, instead of using cropped small patches in some other works (Çiçek, Abdulkadir, Lienkamp, Brox & Ronneberger 2016) (Yu, Yang, Chen, Qin & Heng 2017) for 3D medical image segmentation, our approach uses images with the original size ( $512 \times 512$ ), which maximally fits to the computation capacity of GPUs and keeps the structural context of mandibles.

Second, we used the multi-planar volume-to-slice training strategy which takes into account the structural continuity of mandibles from different views. Naturally, the mandible anatomical structure is very complicated. The CT imaging quality is severely and easily disrupted by dental prosthesis or fillings in the teeth since they are produced by highly attenuated materials. These materials lead to noisy and ambiguous boundaries in the CT imaging. In order to overcome this shortcoming, we used multi-planar volume-to-slice strategy for training, which can learn the 'upper and lower structure' of the mandible in each CT slice. To increase the robustness of the segmentation method, we only used the original CT data without any pre-processing step.

The achieved experimental results demonstrate the feasibility and effectiveness of the proposed approach in 3D mandible segmentation and could also be applied to other segmentation tasks. The current implementation treats the corresponding 3D CT image segmentation of the mandible as a video object segmentation with continuous structure, in which we consider the special continuous structure of the mandible. Therefore, the technique of taking into account the adjacent slices (frames) information could be transferred to the application of object segmentation in video frames. Moreover, the collection of contextual and shape information from different planes assures the



**Figure 10.** Insight of learned U-Net in Axial plane. (a) The input CT image. (b)-(g) Intermediate output of the 21-st layer. All the intermediate outputs are randomly chosen from the 21-st layer.

sufficient extraction of useful information from input images, which could provide a future research direction for 3D image segmentation. Finally, the proposed work could be applied to other organ segmentation or in other imaging modalities.

Despite the promising results, this study also has several limitations. First, in this experiment, we use 109 head and neck scans for the training and the validation of the proposed approach. Our data set may be limited and cannot sufficiently represent the general population in clinical practice. The experimental results need to be validated with large variety data sets in future studies. Second, the inconsistency of the manual annotations on the inferior and superior teeth influences the performance of the proposed approach as shown in the examples on top of Figure 7-9. This could be improved by setting the manual annotation protocol to keep consistency. Third, the accurate evaluation of mandible segmentation remains a challenge and is subjective due to the inconsistency in manual segmentation for the ambiguous boundaries in the parts of teeth and condyles of the mandible. Thus, developing more accurate mandible segmentation and evaluation methods is still one of the research focuses. Further evaluation of our approach is required to assess its performance in clinical practice. This could be done through more massive and intensive experiments in or outside our maxillofacial oncology center. More experimental data sets should be used to verify this technique in future studies. Besides, it is of great importance to verify the automatic mandible segmentation in the 3D virtual planning of craniofacial tumor resection and free flap reconstruction.

**5. Conclusion**

This study proposes an end-to-end approach for automatic segmentation of the mandible from head and neck CT scans. Our approach takes into account the input images from three planes for 2D segmentation of mandibles in multi-planar CT slices and then combines the resulting 2D segmentation into 3D segmentation. We implement the proposed approach on 109 head and neck CT scans and the PDDCA public dataset. We achieve an average Dice coefficient of 0.881 and an average surface error of 0.5791 mm on 109 CT scans and an average Dice coefficient of 0.9328 and 95HD of 1.4333 mm on PDDCA dataset. The experimental results demonstrate the effectiveness of the proposed approach in mandible segmentation and its potential employment in 3D virtual planning of craniofacial tumor resection and free flap reconstruction.

*Acknowledgments*

The author is supported by a joint PhD fellowship from China Scholarship Council (CSC 201708440222). The authors acknowledge Erhan Saatcioglu for the training data preparation. We also thank Dr. E.J.K. Noach for proofreading the manuscript.

## References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M. et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467*.
- Abdi, A. H., Kasaei, S. & Mehdizadeh, M. (2015). Automatic segmentation of mandible in panoramic x-ray, *Journal of Medical Imaging* **2**(4): 044003.
- Albrecht, T., Gass, T., Langguth, C. & Lüthi, M. (2015). Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning, *Head and Neck Auto-Segmentation Challenge (MICCAI), Munich*.
- Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE transactions on pattern analysis and machine intelligence* **39**(12): 2481–2495.
- Bankman, I. (2008). *Handbook of medical image processing and analysis*, Elsevier.
- Bittermann, G., Scheifele, C., Prokic, V., Bhatt, V., Henke, M., Grosu, A.-L., Schmelzeisen, R. & Metzger, M. C. (2013). Description of a method: computer generated virtual model for accurate localisation of tumour margins, standardised resection, and planning of radiation treatment in head & neck cancer surgery, *Journal of Cranio-Maxillofacial Surgery* **41**(4): 279–281.
- Blaschke, T., Burnett, C. & Pekkarinen, A. (2004). Image segmentation methods for object-based analysis and classification, *Remote sensing image analysis: Including the spatial domain*, Springer, pp. 211–236.
- Chen, A. & Dawant, B. (2015). A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images, *Head and Neck Auto-Segmentation Challenge (MICCAI), Munich*.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE transactions on pattern analysis and machine intelligence* **40**(4): 834–848.
- Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation, *arXiv preprint arXiv:1706.05587*.
- Chollet, F. et al. (2015). Keras, <https://keras.io>.
- Chuang, Y. J., Doherty, B. M., Adluru, N., Chung, M. K. & Vorperian, H. K. (2017). A novel registration-based semiautomatic mandible segmentation pipeline using computed tomography images to study mandibular development, *Journal of computer assisted tomography*.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 424–432.
- Dong, C., Loy, C. C., He, K. & Tang, X. (2014). Learning a deep convolutional network for image super-resolution, *European conference on computer vision*, Springer, pp. 184–199.
- Essig, H., Rana, M., Kokemueller, H., von See, C., Ruecker, M., Tavassol, F. & Gellrich, N.-C. (2011). Pre-operative planning for mandibular reconstruction-a full digital planning workflow resulting in a patient specific reconstruction, *Head & neck oncology* **3**(1): 45.
- Fu, H. (2012). *Semantic image understanding: from pixel to word*, PhD thesis, University of Nottingham.
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V. & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation, *arXiv preprint arXiv:1704.06857*.
- Ghafoorian, M., Karssemeijer, N., Heskes, T., Uden, I. W., Sanchez, C. I., Litjens, G., Leeuw, F.-E., Ginneken, B., Marchiori, E. & Platel, B. (2017). Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities, *Scientific Reports* **7**(1): 5110.
- Gollmer, S. T. & Buzug, T. M. (2012). Fully automatic shape constrained mandible segmentation from cone-beam ct data, *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium*

- on, IEEE, pp. 1272–1275.
- He, J., Yang, Y., Wang, Y., Zeng, D., Bian, Z., Zhang, H., Sun, J., Xu, Z. & Ma, J. (2019). Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction, *IEEE transactions on medical imaging* **38**(2): 371–382.
- Huff, T. J., Ludwig, P. E. & Zuniga, J. M. (2018). The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning, *Expert review of medical devices* **15**(5): 349–356.
- Ibragimov, B. & Xing, L. (2017). Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks, *Medical physics* **44**(2): 547–557.
- Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv preprint arXiv:1502.03167*.
- Ker, J., Wang, L., Rao, J. & Lim, T. (2018). Deep learning applications in medical image analysis, *Ieee Access* **6**: 9375–9389.
- Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.
- Kraeima, J., Schepers, R. H., van Ooijen, P. M., Steenbakkers, R. J., Roodenburg, J. L. & Witjes, M. J. (2015). Integration of oncologic margins in three-dimensional virtual planning for head and neck surgery, including a validation of the software pathway, *Journal of Cranio-Maxillofacial Surgery* **43**(8): 1374–1379.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning, *nature* **521**(7553): 436.
- Li, S., Zeng, D., Peng, J., Bian, Z., Zhang, H., Xie, Q., Wang, Y., Liao, Y., Zhang, S., Huang, J. et al. (2019). An efficient iterative cerebral perfusion ct reconstruction via low-rank tensor decomposition with spatial-temporal total variation regularization, *IEEE transactions on medical imaging* **38**(2): 360–370.
- Lin, G., Milan, A., Shen, C. & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Lin, H.-T. & Lin, C.-J. (2003). A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods, *submitted to Neural Computation* **3**: 1–32.
- Long, J., Shelhamer, E. & Darrell, T. (2015). Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- Mannion-Haworth, R., Bowes, M., Ashman, A., Guillard, G., Brett, A. & Vincent, G. (2015). Fully automatic segmentation of head and neck organs using active appearance models, *MIDAS J.*
- Milletari, F., Navab, N. & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation, *3D Vision (3DV), 2016 Fourth International Conference on*, IEEE, pp. 565–571.
- Mortazi, A., Burt, J. & Bagci, U. (2017). Multi-planar deep segmentation networks for cardiac substructures from mri and ct, *International Workshop on Statistical Atlases and Computational Models of the Heart*, Springer, pp. 199–206.
- Nair, V. & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.
- Peng, C., Zhang, X., Yu, G., Luo, G. & Sun, J. (2017). Large kernel matters—improve semantic segmentation by global convolutional network, *arXiv preprint arXiv:1703.02719*.
- Qin, W., Wu, J., Han, F., Yuan, Y., Zhao, W., Ibragimov, B., Gu, J. & Xing, L. (2018). Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation, *Physics in Medicine & Biology* **63**(9): 095017.
- Raudaschl, P. F., Zaffino, P., Sharp, G. C., Spadea, M. F., Chen, A., Dawant, B. M., Albrecht, T., Gass, T., Langguth, C., Lüthi, M. et al. (2017). Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015, *Medical physics* **44**(5): 2020–2036.

- Ren, X., Xiang, L., Nie, D., Shao, Y., Zhang, H., Shen, D. & Wang, Q. (2018). Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images, *Medical physics* **45**(5): 2063–2075.
- Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.
- Schepers, R. H., Raghoobar, G. M., Vissink, A., Stenekes, M. W., Kraeima, J., Roodenburg, J. L., Reintsema, H. & Witjes, M. J. (2015). Accuracy of fibula reconstruction using patient-specific cad/cam reconstruction plates and dental implants: a new modality for functional reconstruction of mandibular defects, *Journal of Cranio-Maxillofacial Surgery* **43**(5): 649–657.
- Shen, D., Wu, G. & Suk, H.-I. (2017). Deep learning in medical image analysis, *Annual review of biomedical engineering* **19**: 221–248.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research* **15**(1): 1929–1958.
- Tong, N., Gou, S., Yang, S., Ruan, D. & Sheng, K. (2018). Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks, *Medical physics* **45**(10): 4558–4567.
- Torosdagli, N., Liberton, D. K., Verma, P., Sincan, M., Lee, J., Pattanaik, S. & Bagci, U. (2017). Robust and fully automated segmentation of mandible from ct scans, *Biomedical Imaging (ISBI 2017)*, *2017 IEEE 14th International Symposium on*, IEEE, pp. 1209–1212.
- Weijs, W. L., Coppen, C., Schreurs, R., Vreeken, R. D., Verhulst, A. C., Merckx, M. A., Bergé, S. J. & Maal, T. J. (2016). Accuracy of virtually 3d planned resection templates in mandibular reconstruction, *Journal of Cranio-Maxillofacial Surgery* **44**(11): 1828–1832.
- Wu, J., Tha, K. K., Xing, L. & Li, R. (2018). Radiomics and radiogenomics for precision radiotherapy, *Journal of radiation research* **59**(suppl\_1): i25–i31.
- Yu, F. & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions, *arXiv preprint arXiv:1511.07122*.
- Yu, L., Yang, X., Chen, H., Qin, J. & Heng, P.-A. (2017). Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images., *AAAI*, pp. 66–72.
- Yuheng, S. & Hao, Y. (2017). Image segmentation algorithms overview, *arXiv preprint arXiv:1707.02051*.
- Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. (2017). Pyramid scene parsing network, *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890.
- Zhu, W., Huang, Y., Tang, H., Qian, Z., Du, N., Fan, W. & Xie, X. (2018). Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation, *arXiv preprint arXiv:1808.05238*.